Journal of Forecasting, Vol. 3, 197–204 (1984)

Improved Methods of Combining Forecasts

CLIVE W. J. GRANGER
RAMU RAMANATHAN
University of California, San Diego, U.S.A.

ABSTRACT

It is well known that a linear combination of forecasts can outperform individual forecasts. The common practice, however, is to obtain a weighted average of forecasts, with the weights adding up to unity. This paper considers three alternative approaches to obtaining linear combinations. It is shown that the best method is to add a constant term and not to constrain the weights to add to unity. These methods are tested with data on forecasts of quarterly hog prices, both within and out of sample. It is demonstrated that the optimum method proposed here is superior to the common practice of letting the weights add up to one.

KEY WORDS Combining ARMA models Econometrics

It is clear that there are many ways to forecast time series. For a given information set forecasts may be optimal or suboptimal, use a linear formulation or some other specific functional form, use time-invariant or time-varying coefficients, be rich in dynamic specifications or use naïve, unsophisticated ones. There are usually also a variety of information sets to be considered. Choices of forecast types can be made on cost considerations, be based on specific economic theories or depend on the abilities of the analyst. If there are available two sets of one-step forecasts from two competing theories, functional forms or information sets, then it has been known for some time that a linear combination of the two forecasts may outperform both of them. In this paper we extend previous discussions about how combinations should be formed, the properties of the combined forecasts and how the results may be interpreted. However, the discussion here is only of linear combinations.

Consider initially the case where there are two unbiased one-step forecasts $f_{n,1}$, $g_{n,1}$ of x_{n+1} made at time n, so that $E[(x_{n+1} - f_{n,1})] = 0$ and similarly for $g_{n,1}$. The combinations considered by Bates and Granger (1969), Nelson (1972), Dickinson (1975) and many other writers in the field are of the form

$$c_{n,1} = \alpha f_{n,1} + (1 - \alpha) g_{n,1} \tag{1}$$

and it was shown in the above references that typically $c_{n,1}$ is a superior forecast, in terms of the mean squared error of the forecast error, than either component, f or g. A combination of type (1) will be called the 'constrained form'. It does have the advantage that if f and g are unbiased then

0277-6693/84/020197-08$01.00

Received July 1982
Revised May 1983

necessarily c will be unbiased, and this is presumably why the constrained form is so widely used. If the error is

$$e_{c,n} = x_{n+1} - c_{n,1}$$

then one can write

$$x_{n+1} - g_{n,1} = \alpha(f_{n,1} - g_{n,1}) + e_{c,n} \tag{2}$$

The coefficient α is then chosen to minimize the MSE of $e_{c,n}$, which with unbiased forecasts is just the variance of the forecast error. Using least squares to estimate α in (2) immediately shows an important problem with the constrained form, as $e_{c,n+1}$ will be uncorrelated with $(f-g)$, by construction, but not necessarily with f and g individually. It is, thus, possible that $e_{c,n+1}$ can be forecast from f_n, g_n and so the combination is not optimal. However, merely moving to an unconstrained form loses the unbiasedness property. A simple solution is to introduce a further unbiased forecast of x_{n+1}, its unconditional mean m, even though this will be generally a poor forecast. Thus, the extended, constrained combination is

$$c^*_{n,1} = \alpha_1 f_{n,1} + \alpha_2 g_{n,1} + \alpha_3 m \tag{3}$$

where $m = E[x_{n+1}]$, assuming x_t to be a stationary series, and

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

The obvious way to proceed is to perform the least-squares regression

$$x_{n+1} = \alpha_1 f_{n,1} + \alpha_2 g_{n,1} + a + e^*_{c,n+1}$$

where a is a constant and α_1, α_2 are unconstrained. By construction $\mathrm{corr}(f_{n,1}, e^*_{c,n+1}) = 0$ and similarly with $g_{n,1}$, so that at least the information in (f, g) is being fully used in a linear fashion. Equation (3) is called the 'unconstrained form'. The properties of the two forms are considered analytically in the next section and they are compared empirically in Section 2. The unconstrained form is obviously particularly relevant when the component forecasts are biased.

The combination of forecasts can be considered on the simple pragmatic level of arriving at better forecasts but, in some cases, a deeper interpretation can be given. If f and g are based on the same information set, then when $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, it can be concluded that neither is optimal. For example, the two forecasts may be based on different assumptions about the appropriate functional form, linear and log-linear obviously come to mind and then, if the combination successfully beats both components, it shows that the best functional form is neither of those originally selected.

Now consider the case of two different information sets. Theory A (say 'monetarist') may suggest that information set $I_{A,n}$ is appropriate whereas a competing theory B ('Keynesian') may suggest set $I_{B,n}$. Both models can produce forecasts of x_{n+1} and, if they are optimal according to a least-squares criterion, the resulting forecasts will be

$$f_{n,1} = E[x_{n+1} | I_{A,n}]$$
$$g_{n,1} = E[x_{n+1} | I_{B,n}]$$

The optimal forecast based on all possible information I_n, which is the union of $I_{A,n}$ and $I_{B,n}$, is then

$$h_{n,1} = E[x_{n+1} | I_n]$$

However, to consider this complete case may be very expensive. A simple test to see if this extra effort is likely to be worth while, is to consider the particular set of forecasts based on I_n

$$c_{n,1} = \alpha_1 f_{n,1} + \alpha_2 g_{n,2} \tag{4}$$

where α_1, α_2 are chosen so that $E[(x_{n+1} - c_n)^2]$ is minimized. If $f_{n,1}$, say, is optimal with respect to I_n then necessarily $\alpha_1 = 1, \alpha_2 = 0$ but if neither f or g is optimal then neither α_1, α_2 will be zero. There is no necessity that $c_{n,1}$ will be the same as $h_{n,1}$. Thus, an indication can be given that neither initial model is optimal but the combination process achieved will generally not be the optimum, although pragmatically a superior forecast is immediately found.

Throughout, the criterion used to distinguish between forecasts is mean squared error, based on the usual minimization of a squared error cost function. Using a different cost function could lead to different results, e.g. Bordley (1982). However, as there is little practical evidence on actual cost functions appropriate to users of forecasts, and given the general results of Granger (1969) on the approximation of any cost function by a squared error form, we feel that no further justification for our assumption is required.

1. THE THEORY OF COMBINING

In this section three different combining procedures are considered, involving various possible restrictions. A single series x_{t+1} is to be forecast using combinations of k forecasts, $f_{j,t}$, for the period $t = 0, \ldots, n-1$. The notation used is

$x^T = (x_1, x_2, \ldots, x_n)$ is a $1 \times n$ vector of values of x to be forecast, so that the series $x_t, t = 1, \ldots, n$ is made into a vector

$f_j^T = (f_{j0}, f_{j1}, \ldots, f_{j,n-1})$ is a $1 \times n$ vector of forecasts from the jth model made at different times

$F = (f_1, f_2, f_3, \ldots, f_k)$ is an $n \times k$ matrix of forecast values

$l =$ is a vector of 1s of appropriate dimension

Method A

Let $F\alpha$ be the unrestricted combined forecast, where α is a $k \times 1$ vector of weights for the f_js. The forecast error is $e_A = x - F\alpha$. Suppose that α is determined so as to minimize the sum of squared errors of forecasts. That is, we minimize $(x - F\alpha)^T(x - F\alpha)$ with respect to α. The solution is given by

$$F^T(x - F\alpha) = 0 \qquad \text{or} \qquad \hat{\alpha} = (F^T F)^{-1} F^T x \tag{5}$$

The combined forecast is $\hat{x}_A = F\hat{\alpha} = F(F^T F)^{-1} F^T x$. The attained minimum sum of squared forecast errors is

$$Q_A = (x - \hat{x}_A)^T (x - \hat{x}_A) = x^T x - x^T F\hat{\alpha}$$

It will readily be noted that this is nothing but regressing x against $f_1, f_2, \ldots, f_k$ with *no constant term*. The value of Q_A is simply the sum of squared residuals available in every regression program. The problem here is that the forecast errors need not average to zero. These forecasts will, thus, be biased in general. The forecast error vector is $\hat{e}_A = x - F\hat{\alpha}$. There is no reason why $l^T\hat{e}_A = l^T(x - F\hat{\alpha})$ should be zero. This is true even if each f_j has zero mean error forecast. Suppose $l^T(x - f_j) = 0$ for every j, that is every individual forecast is unbiased. In matrix notation, this means $(l^T x) l^T = l^T F$. Hence, $(l^T x) l^T \hat{\alpha} = l^T F\hat{\alpha}$. For the combined forecast $F\hat{\alpha}$ to have zero error mean, we need $l^T F\alpha = l^T x$. This means from the above that $(l^T x) l^T \hat{\alpha} = l^T x$. If, as is very likely, $l^T x \neq 0$ (i.e. x does not have a zero mean), then the combined forecast will be unbiased only when $l^T\hat{\alpha} = 1$, *that is, when the weights add up to unity.* Thus, a pair of sufficient conditions for zero combined forecast bias are:

(a) each forecast f_j has zero error mean ($l^T x = l^T f_j$)
(b) the weights add up to 1 ($l^T\hat{\alpha} = 1$).

In practice, there is no reason why either condition will hold. In particular, condition (b) will be violated unless imposed *a priori*.

Method B

Now consider the case in which the weights are constrained to sum to unity, so that $(x - F\beta)^{\mathrm{T}}(x - F\beta)$ is minimized subject to the restriction $l^{\mathrm{T}}\beta = 1, \beta$ being the weights obtained here. Considering

$$\min_{\beta} (x - F\beta)^{\mathrm{T}}(x - F\beta) + 2\lambda_{\mathrm{B}}(l^{\mathrm{T}}\beta - 1)$$

where λ_{B} is a Lagrangian multiplier, the first-order condition is

$$F^{\mathrm{T}}(x - F\beta) - \lambda_{\mathrm{B}} l = 0$$

$$\hat{\beta} = (F^{\mathrm{T}}F)^{-1}F^{\mathrm{T}}x - \lambda_{\mathrm{B}}(F^{\mathrm{T}}F)^{-1}l = \hat{\alpha} - \lambda_{\mathrm{B}}(F^{\mathrm{T}}F)^{-1}l \tag{6}$$

using equation (5). The unknown λ_{B} is obtained from the restriction $l^{\mathrm{T}}\beta = 1$, so that

$$\lambda_{\mathrm{B}} = (l^{\mathrm{T}}\alpha - 1)/[l^{\mathrm{T}}(F^{\mathrm{T}}F)^{-1}l] \tag{7}$$

The minimum value of the sum of squared residuals is

$$Q_{\mathrm{B}} = Q_{\mathrm{A}} + \lambda_{\mathrm{B}}^2[l^{\mathrm{T}}(F^{\mathrm{T}}F)^{-1}l] \tag{8}$$

Evidently, $Q_{\mathrm{B}} \geq Q_{\mathrm{A}}$ and, hence, there is a loss in mean square error due to the constraint $l^{\mathrm{T}}\beta = 1$. Computationally, method B is equivalent to regressing $(x - f_k)$ against $(f_1 - f_k)$, $(f_2 - f_k), \ldots$, $(f_{k-1} - f_k)$. The weight for f_k is simply 1 − (sum of weights of $f_1, f_2, \ldots, f_{k-1}$). This regression also is without a constant. The sum of squared residuals obtained from a standard computer program is Q_{B}. The gain in mean squared error from relaxing the condition that the forecast weights add up to unity is then easily obtained.

From Condition (a) of Method A, the combined forecast is unbiased when each individual forecast is unbiased. However, if $l^{\mathrm{T}}x \neq l^{\mathrm{T}}f_j$ for some j, then there is no guarantee that the combined forecast will have a zero mean error.

As mentioned in the previous section, there is nothing sacred about the weights adding up to unity, although that seems to be the common practice. Furthermore, there is no reason to believe that every alternative forecast will be unbiased. It can be asked whether such possibly biased forecasts can be combined to yield a forecast with zero average error. The answer is yes and this point is explored in detail in the next method.

Method C

The third method of combining has no restrictions on the weights, but a constant term is added, which requires consideration of

$$\min (x - \delta_0 l - F\delta)^{\mathrm{T}}(x - \delta_0 l - F\delta)$$

with no restrictions, δ_0 being the constant term and δ being the weights for the k forecasts. The normal equations are

$$F^{\mathrm{T}}(x - \delta_0 l - F\delta) = 0 \qquad \text{and} \qquad l^{\mathrm{T}}(x - \delta_0 l - F\delta) = 0$$

and these are solved by

$$\hat{\delta} = \hat{\alpha} - \hat{\delta}_0(F^{\mathrm{T}}F)^{-1}F^{\mathrm{T}}l \tag{9}$$

$$\hat{\delta}_0 = (l^{\mathrm{T}}x - l^{\mathrm{T}}F\hat{\delta})/n = l^{\mathrm{T}}\hat{e}_{\mathrm{A}}/(n - \theta) \tag{10}$$

where $\hat{e}_{\mathrm{A}}$ is the vector of errors using method A and $\theta = l^{\mathrm{T}}F(F^{\mathrm{T}}F)^{-1}F^{\mathrm{T}}l$.

The combined forecast obtained here is $\hat{x}_C = \hat{\delta}_0 l + F\hat{\delta}$. Hence, the mean forecast error is

$$l^T(x - \hat{x}_C)/n = (l^T x)/n - (l^T F\hat{\delta})/\eta - \hat{\delta}_0 = 0$$

from (10). Noting that

$$\hat{e}_C = x - \hat{x}_C = x - \hat{\delta}_0 l - F\hat{\delta} = \hat{e}_A - \hat{\delta}_0[I - F(F^T F)^{-1}F^T]l$$

it follows that the attained sum of squared errors is

$$Q_C = Q_A - 2\hat{\delta}_0 l^T[I - F(F^T F)^{-1}F^T]\hat{e}_A + \hat{\delta}_0^2 l^T[I - F(F^T F)^{-1}F^T]l$$

Because $F^T\hat{e}_A = 0$, this can be reduced to

$$Q_C = Q_A - \frac{(l^T\hat{e}_A)^2}{n - \theta} \leq Q_A \tag{11}$$

where $\theta = l^T F(F^T F)^{-1}F^T l^T$ and $\hat{e}_A$ is the vector of errors using Method A. Method C is clearly the best because it gives the smallest mean squared error and has an unbiased combined forecast *even if individual forecasts are biased.* The common practice of obtaining a weighted average of alternative forecasts should, therefore, be abandoned in favour of an unrestricted linear combination *including a constant term.*

2. AN APPLICATION

Bessler and Brandt (1981) considered the combination of forecasts for quarterly hog prices from an econometric model, an ARIMA model and from expert opinions, for the period 1976-01 to 1979-02. We are grateful to these authors for supplying further terms in these series up to 1981-04, giving a final series of 24 forecasts. These data were analysed in two ways. Initially, the whole series was used to form optimal combining weights, using the three methods discussed in the previous section, combinations in pairs and all three forecasting methods, being compared. The second method used weights derived just for the first 16 terms and the relative forecasting ability then compared out of sample over the remaining eight terms.

Exhibit 1 shows the results of the first exercise, with weight selection and evaluation all occurring in sample. It is seen that the original forecast methods produced somewhat biased forecasts, although the contribution of these biases to the total sum of squared errors is relatively small. The ARIMA forecasts are the best individually, having an MSE roughly 20 per cent better than the expert opinions, which are second best. It is seen that any kind of combining produces spectacular improvements, that Methods A and B lead to biased combinations, but Method C is unbiased. Combination of three forecasts is better than any combination of a pair, although the improvement is sometimes small. Further, as the theory predicted, Method C is consistently better, in terms of mean squared error, than either Methods A or B. It is interesting to note that Method A, which is unconstrained, has weights adding nearly to one, but this is not true for Method C, where the weights on the forecasts add to substantially less than one.

Exhibit 2 shows the weights determined from the first 16 terms and the MSE for the forecasts in the next eight terms using the previously determined weights. This is a much better evaluation procedure, although the short post-sample period means that the results can be misleading owing to sampling error. Mean errors are no longer zero, of course. Method C is consistently superior to the other methods, but now combining three forecasts is not always superior to combining just a pair.

The results in this example certainly indicate the superiority of Method C, although the amount of improvement is not always very large.

Forecast	Mean error	Sum of squared errors	Weights for			
			Const.	Econ.	ARIMA	Expert
Original						
Econometric	−1.71	610.4	—	1.00	—	—
ARIMA	−0.03	420.7	—	—	1.00	—
Expert opinion	0.59	522.7	—	—	—	1.00
Combined method A (unconstrained, no constant term)						
All three	0.06	331.4	0.00	0.35	0.22	0.43
Econ. & ARIMA	0.11	403.4	0.00	0.26	0.73	0.00
ARIMA & expert	0.14	360.7	0.00	0.00	0.62	0.38
Expert & Econ.	0.06	337.4	0.00	0.51	0.00	0.48
Combined method B (no constant, weights sum to 1)						
All three	−0.26	334.7	0.00	0.30	0.27	0.43
Econ. & ARIMA	−0.35	409.8	0.00	0.19	0.81	0.00
ARIMA & expert	0.21	360.8	0.00	0.00	0.45	0.55
Expert & Econ.	−0.44	344.6	0.00	0.62	0.00	0.38
Combined method C (unconstrained with constant)						
All three	0.00	319.6	7.57	0.19	0.26	0.38
Econ. & ARIMA	0.00	372.6	11.80	0.03	0.70	0.00
ARIMA & expert	0.00	325.4	10.65	0.00	0.42	0.34
Expert & Econ.	0.00	327.8	6.80	0.36	0.00	0.48

Sample size = 24.

Exhibit 1. Weights and forecast errors—all in sample

Forecast	Mean error	Sum of squared errors	Weights for			
			Const.	Econ.	ARIMA	Expert
Original						
Econometric	−0.95	322.8	—	1.00	—	—
ARIMA	0.78	245.1	—	—	1.00	—
Expert opinion	−2.13	160.2	—	—	—	1.00
Combined method A (unconstrained, no constant term)						
All three	−0.59	199.8	0.00	0.50	0.16	0.33
Econ. & ARIMA	1.16	246.1	0.00	0.30	0.68	0.00
ARIMA & expert	0.56	217.3	0.00	0.00	0.86	0.14
Expert & Econ.	−0.94	205.0	0.00	0.59	0.00	0.40
Combined method B (no constant, weights sum to 1)						
All three	−1.14	199.1	0.00	0.47	0.15	0.38
Econ. & ARIMA	0.51	238.6	0.00	0.16	0.84	0.00
ARIMA & expert	0.32	212.2	0.00	0.00	0.84	0.16
Expert & Econ.	−1.47	206.6	0.00	0.55	0.00	0.45
Combined method C (unconstrained with constant)						
All three	−0.86	193.4	3.50	0.45	0.13	0.34
Econ. & ARIMA	0.96	233.5	2.89	0.25	0.66	0.00
ARIMA & expert	−0.32	180.2	7.72	0.00	0.63	0.20
Expert & Econ.	−1.17	198.8	3.79	0.51	0.00	0.39

Sample size = 16.
Post sample = 8.

Exhibit 2. Out of sample forecast errors

3. CONCLUSION

There are clear and obvious advantages in combining forecasts, both to better understand the generating mechanism of the series and also to pragmatically achieve better forecasts. This paper has discussed how combinations should be formed and has presented evidence that the constrained combination often used is suboptimal.

A number of potential difficulties still remain. Even if a pair of forecasts is combined, and each has white noise forecast errors, there is no reason to suppose that the combination will have white noise errors. Thus, forecasting performance can be improved further by allowing for a temporal structure of the residuals of the combination. This can be done either by using a Box–Jenkins univariate modelling procedure on the residuals or by introducing multi-step forecasts from the original models and combining over all forecasts. Thus, if x_{n+1} is to be forecast, model 1 would produce forecasts $f_{n+1-k,k}$, $k = 1, 2, \ldots, q$ and Method 2 would produce forecasts $g_{n+1-k,k}$, $k = 1, 2, \ldots, q$ and, if these are all combined, the residuals would necessarily have zero autocorrelations at least up to order q. The second technique may involve more effort, but it does better utilize the total information set available.

The original paper on combining forecasts, (Bates and Granger, 1969), emphasized the possibility of the relevant weights changing through time, as would be appropriate if one set of forecasters learned quicker than another or as one model became a better approximation of reality than another, say. Subsequent work has concentrated on time invariant weights, but the potential in using changing weights deserves further consideration. A start has been made by Engle, Granger and Kraft (1982).

A further topic that requires some consideration is when the series being forecast is an integrated progress, so that the change series is stationary. In this case, the levels may not have a mean, in theory, and so adding a constant to the combination, as suggested in Method C of Section 1 above, is no longer appropriate. In this case forecasts of changes should be used in the combination.

ACKNOWLEDGEMENTS

We thank D. A. Bessler and J. A. Brandt for providing us with all the series needed to carry out the empirical analysis of Section 2. We are also indebted to David Lilien for the use of his Micro TSP on an Apple Computer, and to the referees of this journal.

REFERENCES

Bates, J. M. and Granger, C. W. J., 'The Combination of forecasts', *Operations Research Quarterly*, **20** (1969), 451–468.

Bessler, D. A. and Brandy, J. A., 'Forecasting livestock prices with individual and composite methods', *Applied Economics*, **13** (1981), 513–522.

Bordley, R. F., 'The combination of forecasts: a Bayesian approach', *J. Operations Research Society*, **33** (1982), 171–174.

Dickinson, J. P. 'Some statistical results in the combination of forecasts', *Operational Research Quarterly*, **24** (1975), 253–260.

Engle, R. F., Granger, C. W. J. and Kraft, D., 'Combining competing forecasts of inflation using a bivariate ARCH model', Working paper, UCSD Economics Dept, 1982.

Granger, C. W. J., 'Prediction with a generalized cost of error function', *Operations Research Quarterly*, **20** (1969), 199–207.

Granger, C. W. J. and Newbold, P., 'Forecasting transformed series, *J. Royal Stat. Soc. B*, **38** (1976), 189–203.
Nelson, C. R., 'The predictive performance of the F.R.B.–M.I.T.–PENN model of the U.S. economy', *American Economic Review*, **62** (1972), 902–917.

Authors' biographies:
Clive W. J. Granger obtained his Ph.D. from Nottingham in 1959 where later he held the Chair of Applied Statistics and Econometrics. He was a Harkness Fellow at Princeton in 1959–1960 and has also been a visitor at Stanford, Vienna, Canberra and elsewhere. He has published seven books and over 90 articles on forecasting, time series analysis, spectral analysis, econometrics, speculative markets and pricing theory. He has been Professor of Economics, University of California, San Diego since 1976 and is currently Chair of Economics. He is also Fellow of the Econometric Society and associate editor of several journals, including *Applied Economics*, *Energy*, *Journal of Econometrics*, *Journal of the American Statistical Association*, *Business and Economic Statistics* and *Journal of Financial Economics*.

Ramu Ramanathan obtained an M.A. in Mathematics and Statistics from the Indian Statistical Institute and a Ph.D. in Economics from the University of Minnesota. His interests are applied econometrics, economic growth and international trade. He has written an advanced book on growth theory and published numerous articles which have appeared in leading journals in the U.S.A., U.K., Australia and India.

Authors' addresses:
Clive W. J. Granger and Ramu Ramanathan, Department of Economics, University of California, San Diego, La Jolla, California 92093, U.S.A.